

Positive deconvolution for superimposed extended source and point sources.

J.-F. Giovannelli¹ and A. Coulais²

¹ Groupe Problèmes inverses, Laboratoire des Signaux et Systèmes (CNRS – Supélec – UPS), Plateau de Moulon, 91192 Gif-sur-Yvette, France, e-mail: giova@lss.supelec.fr

² Laboratoire d'Étude du Rayonnement de la Matière en Astrophysique (LERMA), Observatoire de Paris, 61 Avenue de l'Observatoire, 75014 Paris, France, e-mail: alain.coulais@obspm.fr

Submitted January, 2004 and revised January, 2005

Abstract. The paper deals with the construction of images from visibilities acquired using aperture synthesis instruments: Fourier synthesis, deconvolution, and spectral interpolation/extrapolation. Its intended application is to specific situations in which the imaged object possesses two superimposed components: (i) an extended component together with (ii) a set of point sources. It is also specifically designed to the case of positive maps, and accounts for a known support. Its originality lies within joint estimation of the two components, coherently with data, properties of each component, positivity and possible support. We approach the subject as an inverse problem within a regularization framework: a regularized least-squares criterion is specifically proposed and the estimated maps are defined as its minimizer. We have investigated several options for the numerical minimization and we propose a new efficient algorithm based on augmented Lagrangian. Evaluation is carried out using simulated and real data (from radio interferometry) demonstrating the capability to accurately separate the two components.

Key words. Techniques: image processing, interferometric, deconvolution, inverse problem, extended/point sources.

1. Introduction

Radio interferometers can be seen as instruments measuring a set of 2D-Fourier coefficients (visibilities) of the brightness distribution of a region in the sky. Visibilities are measured in the Fourier domain (the (u, v) -plane) by means of different baselines (projected distance between cross-correlated antennas). Practically, there are two principal deficiencies (Thompson et al. 2001) in the visibilities

1. the limited coverage of the (u, v) -plane,
2. measurements errors (especially in millimeter range).

Regarding point 1, three limitations are encountered.

- Usually the central part of the aperture (up to the antenna diameter) is not observed. From this stand point, interferometers behave as high pass filters.
- Information above the longest baseline is unavailable. In this sense the instruments behave as low pass filters.
- The (u, v) -plane coverage is irregular, especially when there is a small number of antennas. This results in dirty beam (Fourier transform of visibility weights) with intricate structure and strong sidelobes.

Thus, such instruments can be seen as band pass filters with an intricate impulse response (dirty beam), and noisy output. As a consequence, the available data is relatively poor for imaging objects with various spatial struc-

tures extended over the whole frequency domain. In order to compensate for these deficiencies, a large number of methods (from model fitting to non parametric deconvolution) has been continuously proposed (see review in (Starck et al. 2002)) and specialized for different types of maps. The present paper deals with a particular type of map consisting of the superimposition of two components.

- *Point Source* (PS), or nearly black objects: essentially null-component, with a few strong point sources.
- *Extended Sources* (ES): spatially extended, smooth components.

The problem at hand is to build reliable and accurate estimates of two distinct maps (one for PS, one for ES) from a unique given set of visibilities. The question arises *e.g.* for radio imaging of the solar corona at meter wavelength where very strong storms are superimposed over a more stable and large quiet Sun radio-emission (see Sect. 4.1).

Remark 1. From a statistical standpoint, PS/ES can be modeled as set of uncorrelated/correlated pixels, respectively. In the Fourier plane they are respectively characterized by an extension over the whole frequency domain (PS) and an extension reduced to the low frequencies domain (ES). In particular, both of them have significant components in the low frequencies domain (see Fig. 1-a).

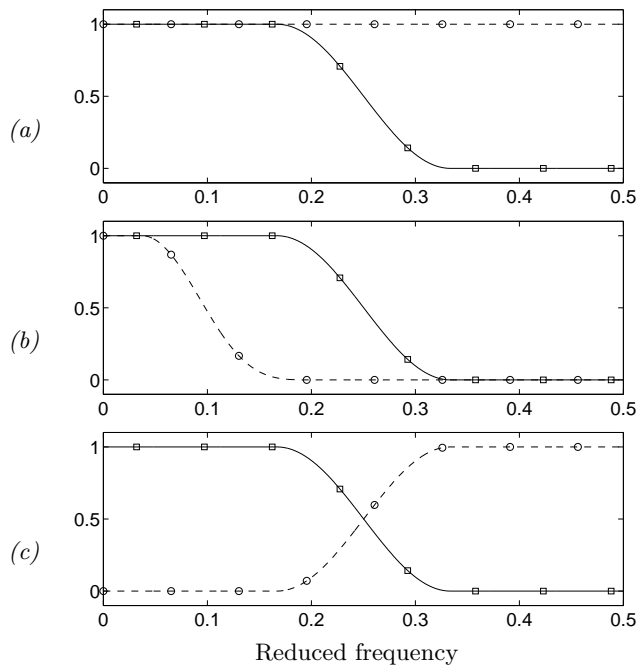


Fig. 1. *a*: The solid line with squares (resp. dashed line with circles) shows spectral content for ES (resp. PS). Both of them have low frequencies components. *b*: The two lines show spectral contents of correlated components (ES), with different level of correlation. *c*: Elementary decomposition for wavelet transform. The solid line with squares (resp. dashed line with circles) shows low (resp. high) frequency content.

1.1. General bibliographical analysis

In order to compensate for the deficiencies in the available data, additional information is (implicitly or explicitly) accounted for. Practically, most existing methods are founded on specific expected properties of observed and reconstructed sources. The proposed analysis relies on underlying decompositions of unknown image.

PS based methods – A first part of existing methods relies on PS properties. Into this category fall original versions of CLEAN (Hogbom 1974; Fomalont 1973), which iteratively withdraw the PS contribution to the dirty map. Early Maximum Entropy Methods (MEM) (Ables 1974) are also founded on the properties of PS: in a regularized context, they introduce separable penalization terms (without pixel interaction) and favor high-amplitude PS.

ES based methods – Two main classes of methods have been proposed to account for the correlation of ES.

- The correlation structure is introduced by a *convolution kernel*. This is the case in MEM with an Intrinsic Correlation Function (ICF) (Gull 1989) and Pixon methods (Dixon et al. 1996; Puetter & Yahil 1999).
- The other class of method relies on *pixel interactive* penalty. The early versions involve quadratic penalties (Tikhonov & Arsenin 1977). Extensions to other

penalties have also been widely developed (O’Sullivan 1995; Snyder et al. 1992; Mugnier et al. 2004).

Mixed ES+SP model – The case of an *explicit model mixing ES and PS* has also been addressed; however, literature in this case is poor. To our knowledge, two papers have been published: (Magain et al. 1998) and (Pirzkal et al. 2000). They introduced the decomposition of the search map as the sum of a PS map and an ES map. From a spectral standpoint, PS / ES are respectively characterized as shown in Fig. 1-a (see also Remark 1). The present paper is founded on this approach (see Sect. 1.2).

Multi-resolution / subband methods – Another class of method received a large attention, namely the multi-resolution and subband approaches.

- The approach proposed by (Weir 1992; Bontekoe et al. 1994) introduces structure by means of different ICF. The unknown map is the sum of several ES, with different level of correlation, *i.e.* several low frequency components. The underlying decomposition is shown in Fig. 1-b in the case of two components.
- We also have witnessed the development of multi-resolution extensions of CLEAN (Wakker & Schwarz 1988) as well as more subtle approaches based on wavelet decomposition and MEM (Starck et al. 1994; Pantin & Starck 1996; Starck et al. 2001). These methods are less specific and widely used for general deconvolution. They aim at reconstructing maps with different scales by splitting the Fourier plane into various zones. They basically rely on (recursive) decomposition in low and high frequencies as shown in Fig. 1-c.

1.2. PS plus ES: proposed developments

As mentioned above, the present paper is devoted to the estimation of two distinct maps (one for ES and one for PS) from a unique set of visibilities. We then naturally resort to the work of Magain et al. (1998) and Pirzkal et al. (2000). In both cases the PS map is written in a parametric manner founded on positions and amplitudes of peaks. Smoothness of the ES is included by means of Gaussian ICF and MEM penalty (Pirzkal et al. 2000) and Tikhonov penalty (Magain et al. 1998). Nevertheless, they both have several limitations. On the one hand, (Pirzkal et al. 2000) relies on the knowledge of the position of the PS which is not available to us. On the other hand, the drawback of (Magain et al. 1998) is twofold.

1. It does not deconvolve with the total PSF.
2. The optimized criterion is intricate w.r.t. the PS positions so, it is not always possible to find the global minimum of the criterion (Magain et al. 1998, p.474).

On the contrary, our approach achieves a complete deconvolution. Moreover, our work introduces properties so that an optimal solution is properly defined and practically attainable. In a unique coherent framework,

the proposed method simultaneously accounts for intricate dirty beam, noise, the existence of point sources superimposed onto a smooth component, positivity, and the possible knowledge of a support. The estimated maps are defined as the constrained minimizer of a penalized least-squares criterion specifically adapted to this situation. So that, the method assigns a coherent value to unmeasured Fourier coefficients. The basic ideas developed here have already been partly presented within spectral analysis (Ciuciu et al. 2001), spectrometry (Mohammad-Djafari et al. 2002) and satellite imaging (Samson et al. 2003).

The paper is organized as follows. In Sect. 2, we define notations and state the problem in three classical forms: Fourier synthesis, spectral extrapolation / interpolation and deconvolution. All three cases concern rank-deficient linear inverse problems with additive noise. The proposed method is presented in Sect. 3. Section 3.1 introduces the regularization principles used in the subsequent sections; Sections 3.2 and 3.3 respectively deal with PS and ES map; Section 3.4 is devoted to the main contribution: the reconstruction of two maps simultaneously, one consisting of PS, and the other of ES. Simulation and real data computations are presented throughout Sect. 4. From a numerical optimization viewpoint, the proposed method reduces to a constrained quadratic programming problem and various options have been studied and compared. The proposed algorithm founded on augmented Lagrangian principle is presented in Sect. 5. In Sect. 6 we set out conclusions and perspectives.

2. Problem statement and least squares solution

The usual model¹ for the instrument writes as a weighted truncated noisy Fourier transform (discrete and regular):

$$\mathbf{y} = \mathbf{W}\mathbf{T}\mathbf{F}\mathbf{x} + \mathbf{b}, \quad (1)$$

where $\mathbf{x} \in \mathbb{R}^N$ is the unknown map and \mathbf{y} and $\mathbf{b} \in \mathbb{C}^M$ are the Fourier coefficient and noise (N unknown parameters for M measurements). \mathbf{F} is the $N \times N$ normalized FFT matrix and \mathbf{T} is a 0/1-binary truncation (or sampling) $M \times N$ matrix (\mathbf{T} discards frequencies outside the (u, v) -plane coverage). \mathbf{W} is a $M \times M$ diagonal matrix accounting for visibility weights. For the sake of simplicity and in accordance with real data processed in Sect. 4.3, the subsequent developments are devoted to unitary weights $\mathbf{W} = \mathbf{I}_M$; they can easily be extended to include non unitary ones. Appendix B gives useful properties of these matrices. The reconstruction of \mathbf{x} from \mathbf{y} , *i.e.* the inversion of (1), is a Fourier synthesis problem.

In formulation (1), the data \mathbf{y} are in the (u, v) -plane while the map \mathbf{x} is in the image plane. Two other statements are usually given: one regarding the (u, v) -plane only and the other the image plane exclusively.

¹ In terms of an usual approximation, after calibration, regridding, ... Moreover, for the sake of readability, equations are given in 1D and computation results are presented in 2D.

1. In the Fourier domain, (1) becomes a simple truncation by an invertible change of variable $\hat{\mathbf{x}} = \mathbf{F}\mathbf{x}$:

$$\mathbf{y} = \mathbf{T}\hat{\mathbf{x}} + \mathbf{b}. \quad (2)$$

Its inversion becomes a problem of extrapolating / interpolating “missing” Fourier coefficients.

2. Furthermore, denoting $\bar{\mathbf{y}} = \mathbf{T}^t \mathbf{y}$ the zero-padded data, and $\hat{\mathbf{y}} = \mathbf{F}^\dagger \bar{\mathbf{y}}$ the dirty map, (1) becomes a convolution

$$\hat{\mathbf{y}} = \mathbf{H}\mathbf{x} + \tilde{\mathbf{b}} \quad (3)$$

where $\mathbf{H} = \mathbf{F}^\dagger \mathbf{T}^t \mathbf{T} \mathbf{F}$ is a (circulant) convolution matrix and $\tilde{\mathbf{b}} = \mathbf{T}^t \mathbf{b}$. (Superscripts “t” and “†” respectively denotes matrix transpose and conjugate-transpose). The instrument response (the dirty beam) is read in any one line of \mathbf{H} , up to a circular shift. The inversion becomes a deconvolution problem.

Remark 2. It should be noted, however, that the correlations of \mathbf{b} and $\tilde{\mathbf{b}}$ differ from one another, and that in this sense, the two problems are not equivalent.

Whichever formulation is envisaged, the tackled problem is a rank-deficient linear inverse problem with additive noise. Indeed, the number of observed Fourier coefficient is far less than the number of pixels ($M \ll N$) and the operators $\mathbf{T}\mathbf{F}$ for (1), \mathbf{T} for (2) or \mathbf{H} for (3) have $N - M$ singular values equal to 0, and M singular values equal to 1. Consequently, the least-squares criterion

$$J^{\text{LS}}(\mathbf{x}) = \|\mathbf{y} - \mathbf{T}\mathbf{F}\mathbf{x}\|^2 \quad (4)$$

possesses an infinite number of minimizers. The dirty map is one such solution since it cancels out J^{LS} , and the other ones are obtained by adding maps with frequency components outside the (u, v) -plane coverage only.

3. Regularization

So, the selection of a unique solution requires *a priori* information on the searched maps to be taken into account. In order to achieve this, we resort to regularization techniques (Idier 2001a; Demoment 1989; Tarantola 1987), allowing diverse types of information to be considered, in order to exclude or avoid non desirable solutions.

3.1. Criterion, penalization and constraints

- **Positivity and support.** This information is naturally encoded by hard constraints for the pixels. Let us note \mathcal{M} , the collection of pixels on the map, \mathcal{S} the collection of pixels on a support, and $\bar{\mathcal{S}}$ its complement in \mathcal{M} .
 - (C_s) : support

$$\forall p \in \bar{\mathcal{S}}, \quad x_p = 0.$$

The proposed method takes into account the knowledge of a support (\mathcal{S} is known and $\mathcal{S} \neq \mathcal{M}$) but remains also valid if $\mathcal{S} = \mathcal{M}$.

- (C_p) : positivity

$$\forall p \in \mathcal{M}, \quad x_p \geq 0.$$

This information is taken to be valid in the following sections of this paper and all reconstructed maps will be positive.

- (C_t) : template

$$\forall p \in \mathcal{M}, \quad t_p^- \leq x_p \leq t_p^+.$$

It is also possible account for a known template but it is not numerically investigated in the paper.

- **Correlation structure.** Here, we are concerned with the *a priori* correlation (ES) or non-correlation (PS) of the searched map. In the image plane, this information is naturally coded by penalization terms $R(\mathbf{x})$, as a sum of potential functions ϕ which addresses the pixels.

- (P_c) : the smooth map (ES) is favored by the introduction of interaction terms between pixels

$$R_c(\mathbf{x}) = \sum_{p \sim q} \phi_c[x_q, x_p], \quad (5)$$

where $p \sim q$ symbolizes neighbor pixels.

- (P_s) : on the other hand, separable terms favor PS

$$R_s(\mathbf{x}) = \sum \phi_s[x_p]. \quad (6)$$

These terms independently shrink the pixels to zero and therefore favor quasi-null maps.

- (P_m) : in the following section, we will also need to penalize the average level of the maps

$$R_m(\mathbf{x}) = \phi_m \left[\sum x_p \right] \quad (7)$$

so as to specifically compensate for the absence of Fourier coefficient at null-frequency.

- (P_d) : it is also possible account for a known default map \bar{x} through a specific penalization term such as

$$R_d(\mathbf{x}) = \sum \phi_d[x_p, \bar{x}_p] \quad (8)$$

but this is not numerically investigated here.

A criterion J is then introduced as a combination of some penalization terms (5)-(8) and the data based one (4) according to the objective: PS component (Sect. 3.2), ES component (Sect. 3.3) and both of them simultaneously (Sect. 3.4). In every case, the solution $\hat{\mathbf{x}}$ is defined as the minimizer of J under constraints C_p and C_s:

$$(\mathcal{P}) \quad \begin{cases} \min J(\mathbf{x}) \\ \text{s.t.} \begin{cases} x_p = 0 & \text{for } p \in \bar{\mathcal{S}} \\ x_p \geq 0 & \text{for } p \in \mathcal{M} \end{cases} \end{cases} \quad (9)$$

that is to say, as the solution of problem (P). One property then becomes crucial to the construction of J :

- (P₁) : J is strictly convex and differentiable.

Indeed, under this hypothesis,

1. the problem (P) possesses a unique solution $\hat{\mathbf{x}}$, which allows the proper definition of the estimated map;
2. the solution in question is continuous with respect to the data and to the tuning parameter values;
3. a broad class of optimization algorithms is available.

As J^{LS} is itself (large sense) convex and differentiable, the property (P₁) can be assured if the potential functions are themselves convex and differentiable. Therefore, we resort to this kind of potential.

Remark 3. Non-convex potentials have been introduced in image reconstruction in the 1980s (Geman & Geman 1984; Blake & Zisserman 1987). As they are richer, they allow a sharper description of the searched images. For example, they can integrate binary variables, allowing contour *detection* to be carried out, at the same time as image reconstruction. As a counterpart, the involved criteria can possess numerous local minima. The computational cost for optimization then increases drastically, and sometimes without guarantee against local minima.

In Sect. 5, several optimization schemes have been investigated within the recommended convex framework. Various iterative algorithms solving (P) are concerned, all of them converging to the unique solution $\hat{\mathbf{x}}$ whatever the initialization. The only question at stake is computation time. An other property of J is therefore crucial.

- (P₂) : J is quadratic and circular-symmetric.

This property allows fast optimization algorithms to be put into practice taking advantage of the FFT algorithm: fast criterion calculations, explicit intermediate solutions, ... Since J^{LS} is itself quadratic and circulant, (P₂) is satisfied if the regularization terms are circulant and the potential functions ϕ are Quadratic (Q) or Linear (L).

Remark 4. Mixed convex potentials, generally quadratic about the origin and linear above a certain threshold, are used in image processing (Bouman & Sauer 1993) and especially in astronomical imaging (Mugnier et al. 2004) in order to *preserve* possible edges. From the optimization strategy stand point, recent works (Idier 2001b; Allain et al. 2004) allow to reduce the convex optimization problem to a partially quadratic one. This would make possible the development of an FFT and Lagrangian based algorithm for our PS+ES problem. We regard these forms as perspectives and we will see that forms Q and L are sufficiently rich and adapted to the envisaged contexts.

3.2. Point Sources – separable linear penalty

This section is devoted to PS: the proposed penalization term is of type (6) where ϕ_s is a potential of \mathbb{R}_+ or \mathbb{R}_+^* onto \mathbb{R} , to be specified.

Classical MEM (Nityananda & Narayan 1982; Narayan & Nityananda 1984; Komesaroff et al. 1981; Gull & Daniell 1978; Gull & Skilling 1984; Bhandari 1978; Le Besnerais et al. 1999) come into play, when, for example, $\phi_s[x] = -\log x$, $\phi_s[x] = x \log x$ or

$\phi_s[x] = -x + \bar{x} + x \log x / \bar{x}$ where \bar{x} is a default map (O’Sullivan 1995; Snyder et al. 1992). They have been widely used in the domain and in image reconstruction (Mohammad-Djafari & Demoment 1988). They have the advantage of ensuring the property (P₁) on \mathbb{R}_+^* , so the problem is properly regularized and (P) possesses a unique solution. They also enjoy the advantage of ensuring (*strict*) positivity, thanks to the presence of an infinite derivative at the origin $\phi'_s(0^+) = -\infty$ (Narayan & Nityananda 1986).

However, these functions prohibit null-pixels and this can be seen as a flaw when the searched maps are largely made up of null-pixels. On the other hand, null-pixels are favored by the introduction of a potential ϕ_s which possesses at its origin (Soussen 2000)

- a minimum value, and
- a strictly positive derivative.

Without loss of generality we set: $\phi_s(0) = 0$ and $\phi'_s(0^+) = 1$, while two possibilities allow property (P₂) to be respected: the form L and the more general form Q.

$$L : \phi_s(x) = x$$

$$Q : \phi_s(x) = \alpha x^2 + x$$

The penalization is then written as:

$$R(\mathbf{x}) = \lambda_s \sum x_p + \varepsilon_s \sum x_p^2. \quad (10)$$

The strict convexity property (P₁) imposes $\varepsilon_s > 0$: the L term ensures a positive derivative at the origin, while the Q term ensures strict convexity.

Remark 5. In order to favor high amplitude peaks, a least penalization function is desirable, *i.e.* $\varepsilon_s = 0$. In this case, it is possible that J remains unimodal or strictly convex, although we have no proof of this. This property could depend on the value of λ_s , on the knowledge and form of the support, on the (u, v) -plane coverage or on the data in each particular case.

3.3. Extended Sources – correlated quadratic penalty

This section is devoted to ES: the penalty term of type (5) introduces interactions between neighboring pixels.

O’Sullivan (1995) proposes the use of an I-divergence: $\phi_c[x, x'] = -x + x' + x \log x / x'$ or an Itakura-Saito distance: $\phi_c[x, x'] = -\log x / x' - 1 + x / x'$ in the symmetrized version. As in the case of Sect. 3.2, these allow property (P₁) and positivity to be ensured. However, they prohibit null-pixels and do not ensure property (P₂).

We resort to classical terms of image processing based on finite differences between neighboring pixels. In the simplest case, first order differences yield

$$\phi_c[x, x'] = \phi_c[x - x']$$

where ϕ_c is a potential of \mathbb{R} onto \mathbb{R} to be specified. In order to effectively favor smooth and correlated maps, and due to reasons of symmetry, ϕ_c is chosen to be minimal in 0

and even. In order to ensure property (P₂), we are led to choose ϕ_c in class Q and to reject class L: $\phi_c(x) = x^2$ and

$$R(\mathbf{x}) = \lambda_c \sum_{p=0}^N [x_{p+1} - x_p]^2$$

with the hypothesis $x_0 = x_N$ in order to ensure circularity.

We are here dealing with early regularization techniques, that appeared in the 1960s (Phillips 1962; Twomey 1963; Tikhonov 1963) and were developed in the mid-1970s in works by Tikhonov & Arsenin (1977) in a continuous context and by Hunt (1977) in a discrete context. They are also related to the well-known Wiener filter.

In this form, the strict convexity condition (P₁) is not respected. Indeed, J^{LS} is not sensitive to constant maps (since null-frequency is not observed) and neither is the regularization term (since it is only a function of the difference between pixels). Several options are available for dealing with this indetermination.

1. Support constraint C_s: as soon as the support constraint is valid, if at least one of the pixels is zero ($\mathcal{S} \neq \mathcal{M}$), J is strictly convex on $\mathbb{R}^{\mathcal{S}}$.
2. In the absence of support information, it is sufficient to penalize the mean of the map by a term such:

$$R_m(\mathbf{x}) = \left[\sum x_p \right]^2.$$

Intuitively, it reduces the mean of the map towards 0 and is counterbalanced by the positivity constraint.

3. It is also possible to penalize the quadratic norm of the map by a term such as that introduced in Sect. 3.2.

The penalization thus reads

$$R(\mathbf{x}) = \lambda_c \sum [x_{p+1} - x_p]^2 + \varepsilon_m \left[\sum x_p \right]^2. \quad (11)$$

Under this form, properties (P₁) and (P₂) are satisfied if ($\varepsilon_m > 0, \lambda_c > 0$) in the case $\mathcal{S} = \mathcal{M}$ and ($\varepsilon_m \geq 0, \lambda_c > 0$) in the case $\mathcal{S} \neq \mathcal{M}$.

3.4. Mixed model

The present paper is devoted to maps composed of both types of component simultaneously: ES and PS. Following (Magain et al. 1998) and (Pirzkal et al. 2000), we introduce two maps \mathbf{x}_e and \mathbf{x}_p which describe each component respectively. The direct model (1) becomes:

$$\mathbf{y} = \mathbf{T}\mathbf{F}(\mathbf{x}_e + \mathbf{x}_p) + \mathbf{b}, \quad (12)$$

and the least-squares term

$$J_{\text{Mix}}^{LS}(\mathbf{x}_e, \mathbf{x}_p) = \|\mathbf{y} - \mathbf{T}\mathbf{F}(\mathbf{x}_e + \mathbf{x}_p)\|^2,$$

where the subscript “Mix” stand for Mixed map. This form raises new indeterminates as it now concerns the estimation of $2N$ variables, still from a single set of M Fourier coefficients. However, it allows the explicit introduction of characteristic information about each map through two adapted regularization terms.

1. A separable term for \mathbf{x}_p , identical to that in Sect. 3.2

$$R_s(\mathbf{x}_p) = \sum x_p(p),$$

minimum at 0 and with a strictly positive derivative.

2. An interaction term between neighboring pixels of the map \mathbf{x}_e , identical to that in Sect. 3.3

$$R_c(\mathbf{x}_e) = \sum [x_e(p+1) - x_e(p)]^2.$$

So as to ensure property (P₁), the same terms as in Sect. 3.2 and Sect. 3.3 are added, and the regularized criterion takes the form:

$$\begin{aligned} J_{\text{Mix}}^{\text{Reg}}(\mathbf{x}_e, \mathbf{x}_p) &= J_{\text{Mix}}^{\text{LS}}(\mathbf{x}_e, \mathbf{x}_p) \\ &+ \lambda_s \sum x_p(p) + \varepsilon_s \sum x_p(p)^2 \\ &+ \lambda_c \sum [x_e(p+1) - x_e(p)]^2 + \varepsilon_m \left[\sum x_e(p) \right]^2, \end{aligned} \quad (13)$$

where superscript “Reg” stands for Regularized. Regularization parameters (hyperparameters) λ_c and λ_s tune the smooth and spiky character of maps \mathbf{x}_e and \mathbf{x}_p .

In this form, properties (P₁) – (P₂) are satisfied if ($\lambda_s \geq 0, \lambda_c > 0, \varepsilon_s > 0$) and $\varepsilon_m > 0$ when $\mathcal{S} \neq \mathcal{M}$ or $\varepsilon_m \geq 0$ when $\mathcal{S} = \mathcal{M}$. The couple of maps ($\widehat{\mathbf{x}}_e, \widehat{\mathbf{x}}_p$) is properly defined as the solution of problem (P) and the next section (Sect. 4) gives the first practical results (simulated and real data processing). Section 5 is devoted to a fast optimization algorithm.

4. Computation results

4.1. Nançay radioheliograph

Radio emission of the Sun at meter wavelength is known since World War II. The Nançay radioheliograph (NRH) is a radio-interferometer dedicated to imaging the solar corona and it monitors the radio burst in solar atmosphere at such wavelengths with high temporal rate, adequate spatial resolution and high dynamic.

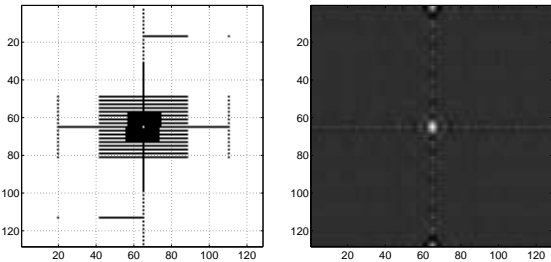


Fig. 2. Left figure shows instantaneous (u, v) -plane coverage (EW array is along vertical direction and NS array is along horizontal direction). Right figure gives the dirty beam, defined as the 2D Fourier transform of the (u, v) -plane coverage with a unitary weight for each visibility.

At such frequencies, mainly two kinds of structures are observed in the corona: (1) larger structures (ES)

and (2) smaller structures (PS). The quiet Sun (1-*i*) is the largest structure, larger than the Sun size in the visible and slowly varying on long term scale (years) (Lantos & Alissandrakis 1996). Medium size structures (1-*ii*) are the radio counterpart of coronal holes and magnetic loops (*plateau*) (Alissandrakis & Lantos 1996), and are also observed simultaneously in soft X rays. The time scale for such structures is days to weeks. They are clearly correlated to persistent structures observed in other wavelength (optical and X-rays) and rotate on the radio maps quasi simultaneously with their optical and X-rays counterparts. The small structures (2) with very high brightness, can often reach several tens of Millions Kelvin (Kerdran & Mercier 1983); they usually have a small life time (few seconds) and are associated to energetic events in the magnetic loops in the Sun’s atmosphere. Correlation with structures observed in other wavelengths is more difficult.

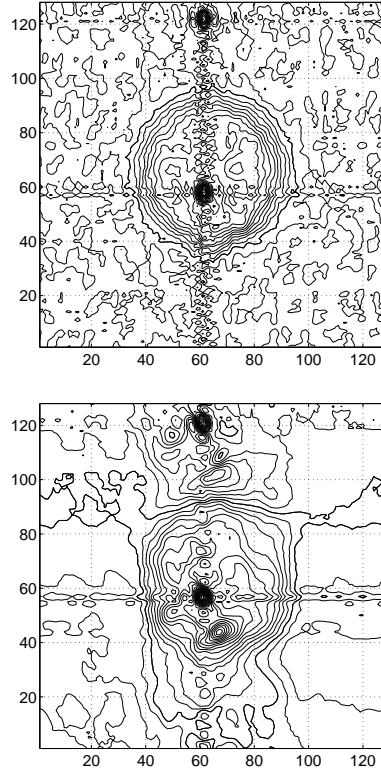


Fig. 3. Dirty maps typically encountered with NRH: simulated data (top) and real data (bottom). Contour levels are -10^{-2} to $5 \cdot 10^{-2}$, step $2.5 \cdot 10^{-4}$ (they are used for all the shown maps).

The NRH is composed by two arrays: one along East-West (EW) direction with 23 antennas, the other along North-South (NS) with 19 antennas. The NRH is operating in the range 150-450 MHz at a time sampling rate of 1/10s., about eight hours a day, with favorable signal to noise ratio. Since the refurbishing of the instrument (Kerdran & Delouis 1997) cross-correlation between most of the antennas in both arrays are available.

As a consequence: (i) 569 non redundant instantaneous visibilities are now available² (with unitary weights) and moreover, (ii) the instantaneous coverage of the (u, v) -plane (shown in Fig. 2) becomes much more uniform. Nevertheless, due to the structure of the arrays, the coverage is not uniform. The central part of the (u, v) -plane essentially consists of two rectangular domains: the central one is a 16×16 square and the larger one is a 32×46 rectangle. With this configuration 2D instantaneous imaging (without Earth rotation aperture synthesis) becomes possible despite strong sidelobes in the dirty beam (see also Fig. 2). As far as the dirty beam is concerned, the maximum value is normalized to 1 and located in the middle of the map at (64, 64). A secondary important lobe partly around (1, 64) and (128, 64) referred to as the aliasing lobe has amplitude 0.70 and characterizes important aliased response. The first negative lobe is -0.10 around the central lobe and -0.23 around the aliasing lobe. Moreover, the first positive lobe is 0.14 near the central lobe and 0.12 near the aliasing lobe. In addition, the FWHM is 4.5 (resp. 4) pixels for the central (resp. aliasing) lobe.

At processed frequency (236 MHz), the field of view³ (FOV) related to the shortest baseline (55 m in NS, 50 m in EW) is $\sim 1^\circ 20'$ and the size of the quiet sun is $\sim 40'$, *i.e.* $\sim 1/2$ FOV. Since at observed frequencies (150-450 MHz) the FWHM of the smallest antenna primary beam (few antennas are 15 m diameter) is much wider than the FOV, a unitary primary beam is appropriate. Moreover, the Shannon criterion is respected if the pixel number is ~ 60 for a FOV of 1° .

With the given characteristics, ES/PS separation must be achieved and reconstruction errors must be as small as possible for both maps, in order to strongly constrain physical models and to monitor position, amplitude and separation of bursts. But imaging the encountered context mixing PS and ES remains difficult and standard methods such as CLEAN and MEM (even in a multiresolution approach) are usually inefficient due to the large background and the intricate mixing of real structures and sidelobes (Coulais 1997). One possible outcome of the present work is to provide to the solar radio community accurate maps from NRH in order to achieve more detailed scientific studies. The following computation study (simulated and real data) is a typical case encountered with NRH and provides a first element in this sense.

4.2. Simulation results

Simulated data

The true ES map \mathbf{x}_e^* (Fig. 4-a), is ranging in amplitude (arbitrary units) from 0 to $5.5 \cdot 10^{-3}$. The true Sun lies in

a disk centered in the middle of the image, *i.e.* (64, 64) with a 64 pixels diameter. The outer part of the disk is zero and the mean of this component is $5.59 \cdot 10^{-4}$. The true PS map \mathbf{x}_p^* (Fig. 4-b) consists of two peaks: the first one is located at (60, 61) with amplitude $5.0 \cdot 10^{-2}$, and the second one overlaps pixels (57, 61) and (57, 62) with respective amplitudes $5.0 \cdot 10^{-2}$ and $4.5 \cdot 10^{-2}$. Data (in the (u, v) -plane) are simulated using the direct model Eq.(12), *i.e.* FFT and truncature, and corrupted by a white, zero-mean complex Gaussian noise with variance $2 \cdot 10^{-7}$. (This noise variance has been chosen in order to mimic real data). The dirty map is shown in Fig. 3. It is clearly dominated by the PS, and the whole map is corrupted by side lobes. Moreover, the two close peaks at location (60, 61) and (57, 61) – (57, 62) are not resolved.

Reconstruction parameters

The supports have been deduced from the dirty map. It is a disk centered at (64, 64) with a 70 pixels diameter for the ES map. Regarding the PS map, the support consists of one disk centered in (58, 61) with a 10 pixels diameter.

In practice, two hyperparameters have to be tuned: λ_c and λ_s (ϵ_s is practically set to 10^{-10}). λ_c must be set in the order of magnitude of eigenvalues of the Hessian of the criterion and is set to $\lambda_c = 2$. λ_s has been empirically selected after several trials in order to visually achieve separation of PS and ES: it has been set to $\lambda_s = 10^{-3}$.

Reconstruction results

Fig. 4 shows the reconstructed maps. A simple qualitative comparison with the references \mathbf{x}_e and \mathbf{x}_p shows that the two components $\widehat{\mathbf{x}}_e$ and $\widehat{\mathbf{x}}_p$ are efficiently separated and accurately reconstructed.

The two peaks of $\widehat{\mathbf{x}}_p$ shown in Fig. 4-d are precisely located at (60, 61) and (57, 61) – (57, 62) (overlapping). The estimated amplitudes are 0.051, 0.048 and 0.043 respectively, *i.e.* an error of less than 5%. Moreover, the two close peaks are separated whereas they are not in the dirty map. This illustrates the resolution capability of the proposed method resulting from both data and accounted information (positivity, support, and PS+ES hypothesis). It is also noticeable that the respective part of flux in overlapped pixels (57, 61) – (57, 62) is correctly restored.

Fig. 4-c gives the estimated ES map $\widehat{\mathbf{x}}_e$. Compared to the true one of Fig. 4-a the main structures are accurately estimated. The contour lines of Fig. 4-c are very similar to the one of Fig. 4-a and the relative reconstruction error is less than 2%. Moreover, the mean of the estimated ES map $\widehat{\mathbf{x}}_e$ is $5.57 \cdot 10^{-4}$ while the true mean is $5.59 \cdot 10^{-4}$: the total flux is correctly estimated. The maximum value is $5.4 \cdot 10^{-3}$ in $\widehat{\mathbf{x}}_e$ whereas it is $5.5 \cdot 10^{-3}$ in \mathbf{x}_e : the dynamic is also correctly retrieved. Nevertheless, a slight distortion located around pixel (65, 60) can be observed in the proposed ES map. It probably results from an imperfect separation of the two components: a slight trace of the dirty beam remains in the estimated ES map. Moreover, the sharp edges of the true Sun are slightly smoothed due to the lack of high frequencies in the available Fourier

² Thanks to Hermitian symmetry 1138 Fourier coefficients are available. The computed (u, v) -plane and map are 128×128 .

³ For a declination 23° and null hour angle (the Sun at noon in summer), the FOV is $87'$ in EW and $86'$ in NS, and the resolution is $3.27'$ in EW and $2.17'$ in NS, since main EW arm is 1600 m with step 50 m and NS arm is 2640 m with step 55 m.

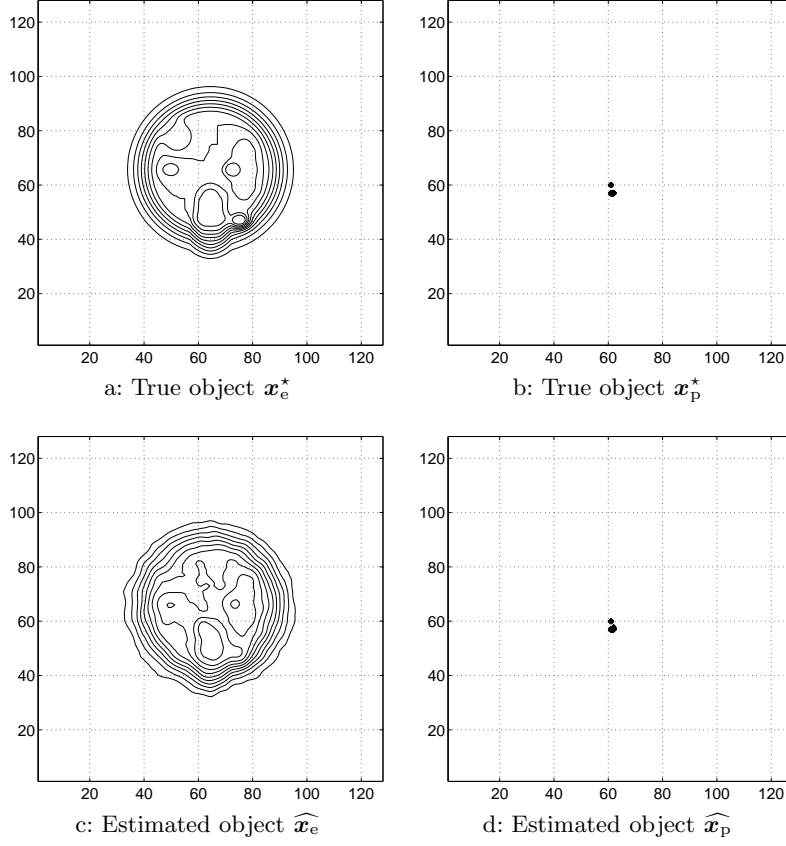


Fig. 4. Simulation results (see Sect. 4.2). Contour levels are the same than in Fig. 3 and 5 for all the maps: the true ES x_e^* (a) and the estimated one \widehat{x}_e (c) as well as for the true PS x_p^* (b) and the estimated one \widehat{x}_p (d).

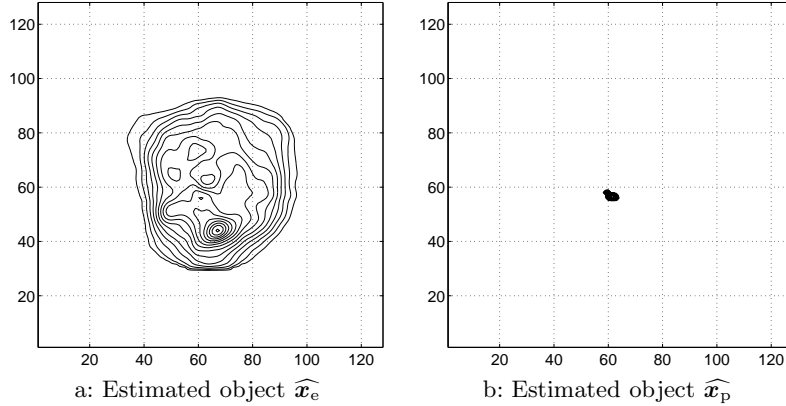


Fig. 5. NRH data processing from typical scientific observation at 236 MHz (see Sect. 4.3). Contour levels are the same than in Fig. 3 and 4. The two components \widehat{x}_e (a) and \widehat{x}_p (b) are clearly separated and deconvolution of both component is clearly achieved. Both maps are positive and the prescribed supports are respected.

coefficients incompletely enhanced by accounted prior information (see Remark 3 and 4).

4.3. Real data computations

This section is devoted to real data processing based on a data set from the NRH⁴. The coverage is identical to the one of simulated data of the previous Section.

The dirty beam is shown in Fig. 2 and the dirty map is shown in Fig. 3. Both dirty beam and dirty map are typically encountered with NRH and are similar to the one simulated in the previous section. As expected, resolution is limited and the quality of the map is entirely contingent upon sidelobes around the brightest point sources (radio burst). Imaging such a complex context mixing PS and ES suffers from intricate mixing of real structures and sidelobes due the brightest ones.

⁴ The eleventh of June, 2004, 13h00, at 236 MHz.

The same supports have been used to compute the real data and the simulated ones. It is a disk centered in the middle of the map with a 70 pixels diameter for the ES map and a disk centered in (58, 61) with 10 pixels diameter for the PS map. The same value of the parameters $\lambda_c = 2$ and $\lambda_s = 10^{-3}$ have been used to compute the real data and the simulated one (ε_s remains set to 10^{-10}).

Estimated maps are shown in Fig. 5-a (ES component) and Fig. 5-b (PS component). The two components $\widehat{\mathbf{x}}_e$ (Fig. 5-a) and $\widehat{\mathbf{x}}_p$ (Fig. 5-b) are clearly separated and both deconvolution is clearly achieved. Both maps are positive and the prescribed supports are respected. Moreover, the $\widehat{\mathbf{x}}_e$ map presents a similar structure to the usual one of the Sun at meter wavelengths without strong point sources (Coulais 1997; Lantos & Alissandrakis 1996).

5. Numerical optimization stage

The estimated maps are defined as the unique solution of the problem (\mathcal{P}) given by (9) which involves the quadratic criterion J given by (13). Up to an additive constant:

$$J(\mathbf{x}) = \frac{1}{2} \mathbf{x}^t \mathbf{Q} \mathbf{x} + \mathbf{q}^t \mathbf{x},$$

where $\mathbf{x} = [\mathbf{x}_e; \mathbf{x}_p]$ collects the two maps (appendix C gives \mathbf{Q} and \mathbf{q}). Thus, (\mathcal{P}) is a convex quadratic program:

$$(\mathcal{P}) \quad \begin{cases} \min \frac{1}{2} \mathbf{x}^t \mathbf{Q} \mathbf{x} + \mathbf{q}^t \mathbf{x} \\ \text{s.t.} \begin{cases} x_p = 0 & \text{for } p \in \bar{\mathcal{S}} \\ x_p \geq 0 & \text{for } p \in \mathcal{M} \end{cases} \end{cases} \quad (14)$$

widely investigated in the optimization literature. The main difficulty is twofold. On the one hand, the non-separability of J together with positivity constraint prevents from explicit optimization. On the other hand, the number of variables is very large. We have investigated most of the proposed methods in the excellent reference book (Nocedal & Wright 2000):

- Constrained gradient
- Gradient projection
- Barrier and interior point
- Relaxation (coordinate-by-coordinate)
- Augmented Lagrangian (method of multipliers)

and have selected the latter as the faster. It is based upon successive optimizations of a Lagrangian function \mathcal{L} founded on Lagrange multipliers ℓ , slack variables s and quadratic penalty. It is computationally based on FFT and threshold, so it is, in addition, very simple to implement.

5.1. Lagrangian Function

The equality constraint $x_p = 0$ ($p \in \bar{\mathcal{S}}$) is introduced by means of a usual Lagrangian term $-\ell_p x_p$ together with a

penalty term $cx_p^2/2$. The entire term write:

$$-\sum_{p \in \bar{\mathcal{S}}} \ell_p x_p + \frac{1}{2} c \sum_{p \in \bar{\mathcal{S}}} x_p^2. \quad (15)$$

The inequality constraint $x_p \geq 0$ ($p \in \mathcal{S}$) is converted into the equality one $s_p - x_p = 0$ using the slack variable $s_p \geq 0$. Lagrange and penalty terms then write:

$$-\sum_{p \in \mathcal{S}} \ell_p (x_p - s_p) + \frac{1}{2} c \sum_{p \in \mathcal{S}} (x_p - s_p)^2. \quad (16)$$

In order to simultaneously process both equality (15) and inequality (16) constraints, we introduce extra slack variables $s_p = 0$ for $p \in \bar{\mathcal{S}}$. The Lagrangian then writes:

$$\mathcal{L}(\mathbf{x}, \mathbf{s}, \boldsymbol{\ell}) = J(\mathbf{x}) - \boldsymbol{\ell}^t (\mathbf{x} - \mathbf{s}) + \frac{1}{2} c (\mathbf{x} - \mathbf{s})^t (\mathbf{x} - \mathbf{s})$$

where \mathbf{s} and $\boldsymbol{\ell}$ collect slack variables s_p and multipliers ℓ_p .

5.2. Algorithm

The algorithm then iterates three steps:

- ① unconstrained minimization of \mathcal{L} w.r.t. \mathbf{x} ,
- ② minimization of \mathcal{L} w.r.t. \mathbf{s} , s.t. $s_p \geq 0$,
- ③ update $\boldsymbol{\ell}$ and c .

The efficiency of the proposed algorithm relies on both slack variables and property (P₂). Roughly speaking, positivity is transferred on slack variables, so, the non separable constrained problem (\mathcal{P}) is split in two subproblems: a non-separable but unconstrained one computable by FFT (step ①) and a constrained but separable one (step ②).

Step ① proceeds by fixing $\boldsymbol{\ell}$ and \mathbf{s} to the current value and then computes the unconstrained minimizer $\tilde{\mathbf{x}}$ of \mathcal{L} . It is an unconstrained convex quadratic problem, so its solution is explicit:

$$\tilde{\mathbf{x}} = -(\mathbf{Q} + c\mathbf{I}_N)^{-1} (\mathbf{q} + [\boldsymbol{\ell} + c\mathbf{s}]),$$

and computable by means of FFT, thanks to circularity.

Step ② updates the slack variables s_p for $p \in \mathcal{S}$ (by construction, $s_p = 0$ for $p \in \bar{\mathcal{S}}$) as the minimizer \tilde{s}_p of \mathcal{L} , subject to $s_p \geq 0$.

$$\tilde{s}_p = \begin{cases} \max(0, cx_p - \ell_p)/c & \text{for } p \in \mathcal{S} \\ 0 & \text{for } p \in \bar{\mathcal{S}} \end{cases}$$

This step is constrained but separable: the constrained minimizer is the unconstrained one if positive and 0 if not.

Step ③ consists in updating the Lagrange multiplier ℓ :

$$\tilde{\ell}_p = \begin{cases} \max(0, \ell_p - cx_p) & \text{for } p \in \mathcal{S} \\ \ell_p - cx_p & \text{for } p \in \bar{\mathcal{S}} \end{cases}$$

This step can also include an update of c (e.g. $\tilde{c} = 1.1c$). Practically, c is not updated (see next subsection).

Steps ① to ③ are iterated until stopping condition is met, e.g. relative variation smaller than 0.1%.

Remark 6. Constrained variables $x_p = 0$ for $p \in \bar{\mathcal{S}}$ can also be eliminated. This is a relevant strategy when using gradient based or relaxation methods. It does not prevent from computing J and its gradient by means of FFT. On the contrary, such a strategy is not relevant here: it would break circularity and prevent from using FFT in step ①.

5.3. Practical case and computations time

This section specializes the algorithm in the case of constant coefficient c . In this case, step ②–③ reduces to:

$$\tilde{\ell}_p + c\tilde{s}_p = \begin{cases} |\ell_p - cx_p| & \text{for } p \in \mathcal{S} \\ \ell_p - cx_p & \text{for } p \in \bar{\mathcal{S}} \end{cases}$$

Moreover, $\mathbf{Q} + c\mathbf{I}_N$ can be inverted once for all and the algorithm then requires 4 FFT per iteration. The algorithm has been used in the previous computations with constant coefficient $c = 10^{-3}$. Convergence is achieved after about 1000 iterations and it takes half a minute⁵.

6. Conclusions

The problem of incomplete Fourier inversion is addressed as it arises in map reconstruction (deconvolution, spectral interpolation/extrapolation, Fourier synthesis). The proposed solution is dedicated to specific situations in which the imaged object involves two components: (i) an extended component together with (ii) a set of point sources. For this cases, new developments are given based on existing work of Magain et al. (1998) and Pirzkal et al. (2000).

The main part of the paper deals with inversion in the regularization framework. It essentially departs from usual strategies by the way it accounts for (1) noise and indeterminacies, (2) smoothness/sharpness prior and (3) positivity and support, in a unique coherent setting. The presented development can also include known template and default map. Thus, a new regularized criterion is introduced and estimated maps are properly defined as its unique minimizer. The criterion is iteratively minimized by means of an efficient algorithm essentially based on Lagrange multiplier which practically requires FFT and threshold only. The minimizer is shown to be both practically reachable and accurate. A first evaluation of the proposed method has been carried out using simulated and real data sets. We demonstrate ability to separate the two components, high resolution capability and high quality of each map. To our knowledge, such a development is an original contribution to the field of deconvolution.

Nevertheless, a further evaluation of the proposed method is desirable. Future work will include systematic evaluation of the capability of the proposed method as a function of (u, v) -plane coverage, PS amplitudes *versus* ES ones, PS position (especially in a subpixelic sense) and

noise level. Such an assessment concerns both simulated and real data. Moreover evaluation of the potentiality of the method on large maps (*e.g.* VLA images), high dynamic imaging (*e.g.* WSRT images) and imaging using millimeter interferometers (*e.g.* IRAM PdBi and ALMA) or optical instruments will be considered.

A part of future work in the field of SP+ES imaging, will include convex non quadratic penalization of ES (see Remark 4). Another part of future work will particularize the proposed method in order to produce maps of ES only and maps of PS only.

A Bayesian interpretation of the proposed method involve truncated Gauss-Markov models (ES component) and exponential white noise (PS component) and formally provides likelihood tools in order to achieve automatic tuning of the hyperparameters. This is a more delicate aspect but it will be addressed in future works.

Acknowledgements. The authors thank Anthony LARUE and Patrick LEGROS for substantial contribution in optimization investigations. The authors are particularly grateful to François VIALLEFOND, Alain Kerdraon, Christophe MARQUÉ, Claude MERCIER and Jérémie LECLÈRE for useful discussions and for providing NRH data. We are grateful to Guy LE BESNERAIS and Éric THIÉBAUT for carefully reading the paper. Special thanks to the incredible, indescribable Grün.

Appendix A: Notations

In this paper, \mathbf{I}_P denotes the $P \times P$ identity matrix and \mathbf{M}^\dagger (resp. \mathbf{M}^t) denotes the complex conjugate transpose (resp. transpose) of a given matrix \mathbf{M} .

Let us note \mathbf{D} the (circulant) first order difference matrix and $\mathbf{\Lambda}_D = \mathbf{FDF}^\dagger$ the diagonalized matrix. Let us also note $\mathbf{1}$ the ones column vector with N components and $\mathring{\mathbf{1}} = \mathbf{F}\mathbf{1}$ its FFT (non-null at null frequency only).

We introduce now two matrices $\mathbf{\Delta}_E$ and $\mathbf{\Delta}_P$ useful to compute ES and PS respectively:

$$\begin{cases} \mathbf{\Delta}_E = \mathbf{\Delta}_C + \lambda_c \mathbf{\Lambda}_D + \varepsilon_m \mathring{\mathbf{1}} \mathring{\mathbf{1}}^\dagger \\ \mathbf{\Delta}_P = \mathbf{\Delta}_C + \varepsilon_s \mathbf{I}_N \end{cases}$$

where $\mathbf{\Delta}_C = \mathbf{T}^t \mathbf{T}$. The three sub-matrices $\mathbf{\Delta}_E$, $\mathbf{\Delta}_P$ and $\mathbf{\Delta}_C$ are diagonal matrices.

Appendix B: Conventions and properties

This Appendix gives several properties of \mathbf{F} and \mathbf{T} introduced in Sect. 2.

- $\mathbf{F}^\dagger \mathbf{F} = \mathbf{F} \mathbf{F}^\dagger = \mathbf{I}_N$: orthonormality of the normalized FFT.
- \mathbf{T} is a truncation operator, $N \times M$, (eliminates coefficient outside the coverage).
- \mathbf{T}^t is a zero-padding operator, $M \times N$, (adds null coefficient outside the coverage).
- $\mathbf{\Delta}_C = \mathbf{T}^t \mathbf{T}$ is a projection matrix, $N \times N$, (nullifies coefficients outside the coverage).
- $\mathbf{T} \mathbf{T}^t = \mathbf{I}_M$.

⁵ Algorithm has been implemented with the computing environments Matlab and IDL on a PC, with a 2 GHz AMD-Athlon CPU, and 512 MB of RAM. Both codes are ~ 50 lines long.

$\rho(\mathbf{x})$	$\hat{\rho}(\hat{\mathbf{x}})$	$\partial\rho/\partial\mathbf{x}$	$\partial\hat{\rho}/\partial\hat{\mathbf{x}}$	$\partial^2\rho/\partial\mathbf{x}^2$	$\partial^2\hat{\rho}/\partial\hat{\mathbf{x}}^2$
$\ \mathbf{y} - \mathbf{T}\mathbf{F}\mathbf{x}\ ^2$	$\ \mathbf{y} - \mathbf{T}\hat{\mathbf{x}}\ ^2$	$-2\mathbf{F}^\dagger\mathbf{T}^\dagger(\mathbf{y} - \mathbf{T}\mathbf{F}\mathbf{x})$	$-2\mathbf{T}^\dagger(\mathbf{y} - \mathbf{T}\hat{\mathbf{x}})$	$2\mathbf{F}^\dagger\mathbf{T}^\dagger\mathbf{T}\mathbf{F}$	$2\mathbf{T}^\dagger\mathbf{T}$
$\mathbf{x}^\dagger\mathbf{D}^\dagger\mathbf{D}\mathbf{x}$	$\hat{\mathbf{x}}^\dagger\mathbf{\Lambda}_D^\dagger\mathbf{\Lambda}_D\hat{\mathbf{x}}$	$2\mathbf{D}^\dagger\mathbf{D}\mathbf{x}$	$2\mathbf{\Lambda}_D^\dagger\mathbf{\Lambda}_D\hat{\mathbf{x}}$	$2\mathbf{D}^\dagger\mathbf{D}$	$2\mathbf{\Lambda}_D^\dagger\mathbf{\Lambda}_D$
$\mathbf{x}^\dagger\mathbf{x}$	$\hat{\mathbf{x}}^\dagger\hat{\mathbf{x}}$	$2\mathbf{x}$	$2\hat{\mathbf{x}}$	$2\mathbf{I}_N$	$2\mathbf{I}_N$
$(\mathbb{1}^\dagger\mathbf{x})^2$	$\hat{\mathbf{x}}(0)^2$	$2\mathbb{1}\mathbb{1}^\dagger\mathbf{x}$	$2\hat{\mathbb{1}}\hat{\mathbb{1}}^\dagger\hat{\mathbf{x}}$	$2\mathbb{1}\mathbb{1}^\dagger$	$2\hat{\mathbb{1}}\hat{\mathbb{1}}^\dagger$
$\mathbb{1}^\dagger\mathbf{x}$	$\hat{\mathbf{x}}(0)$	$\mathbb{1}$	$\hat{\mathbb{1}}$	$\mathbf{0}$	$\mathbf{0}$

(C.1)

Table C.1. Functions, gradients and Hessian of encountered criteria (given as a function of map and their FFT).

Appendix C: Gradient and Hessian calculi

This Appendix is devoted to the vector \mathbf{q} and the matrix \mathbf{Q} involved in the minimized criterion.

\mathbf{q} is a $2N$ components column vector based on the gradient of criterion J , at $\mathbf{x} = 0$. The first part is the dirty map and the second one is the dirty map minus a constant map equal to $\lambda_s/2$. In the Fourier domain, \mathbf{q} reads:

$$\hat{\mathbf{q}} = \left. \frac{\partial J}{\partial \hat{\mathbf{x}}} \right|_{\mathbf{x}=0} = \begin{bmatrix} \frac{\partial J}{\partial \hat{\mathbf{x}}_e} \\ \frac{\partial J}{\partial \hat{\mathbf{x}}_p} \end{bmatrix}_{\mathbf{x}=0} = -2 \begin{bmatrix} \bar{\mathbf{y}} \\ \bar{\mathbf{y}} - \lambda_s \hat{\mathbb{1}}/2 \end{bmatrix}.$$

\mathbf{Q} is a $2N \times 2N$ matrix based on the Hessian of J . The two anti-diagonal elements are the Hessian of the LS term and rely on the dirty beam only. The diagonal elements are the Hessian of J w.r.t. each map \mathbf{x}_p and \mathbf{x}_e . In the Fourier domain, \mathbf{Q} reads:

$$\hat{\mathbf{Q}} = \frac{\partial^2 J}{\partial \hat{\mathbf{x}}^2} = \begin{bmatrix} \frac{\partial^2 J}{\partial \hat{\mathbf{x}}_e^2} & \frac{\partial^2 J}{\partial \hat{\mathbf{x}}_e \partial \hat{\mathbf{x}}_p} \\ \frac{\partial^2 J}{\partial \hat{\mathbf{x}}_p \partial \hat{\mathbf{x}}_e} & \frac{\partial^2 J}{\partial \hat{\mathbf{x}}_p^2} \end{bmatrix} = \begin{bmatrix} \Delta_E & \Delta_C \\ \Delta_C & \Delta_P \end{bmatrix}.$$

Appendix D: Object updates

The present subsection gives details about the step ① of the proposed algorithm (Sect. 5) : the unconstrained minimization of \mathcal{L} w.r.t. \mathbf{x} , *i.e.* the update of \mathbf{x}_e and \mathbf{x}_p . Let us introduce the two vectors

$$\begin{cases} \mathbf{z}_e = \bar{\mathbf{y}} + (\hat{\ell}_e + c \hat{\mathbf{s}}_e)/2 \\ \mathbf{z}_p = \bar{\mathbf{y}} + (\hat{\ell}_p + c \hat{\mathbf{s}}_p)/2 - \lambda_s \hat{\mathbb{1}} \end{cases}$$

based on observed data $\bar{\mathbf{y}}$ and FFT of slack variables and Lagrange multipliers $\hat{\mathbf{s}} = \mathbf{F}\mathbf{s}$ and $\hat{\ell} = \mathbf{F}\ell$ (for each map ES and PS). Let us also introduce two diagonal matrices

$$\begin{cases} \mathbf{M}_E = \Delta_E + c \mathbf{I}_N/2 \\ \mathbf{M}_P = \Delta_P + c \mathbf{I}_N/2 \end{cases}$$

The update reads :

$$\begin{cases} \hat{\mathbf{x}}_e = (\mathbf{M}_E \mathbf{M}_P - \Delta_C^2)^{-1} (\mathbf{M}_P \mathbf{z}_e - \Delta_C \mathbf{z}_p) \\ \hat{\mathbf{x}}_p = (\mathbf{M}_E \mathbf{M}_P - \Delta_C^2)^{-1} (\mathbf{M}_E \mathbf{z}_p - \Delta_C \mathbf{z}_e) \end{cases}$$

easily implemented since $\mathbf{M}_E \mathbf{M}_P - \Delta_C^2$ is diagonal.

References

- Ables. 1974, *Astronomy and Astrophysics Supplement*, 15, 383
- Alissandrakis, C. E. & Lantos, P. 1996, *Solar Physics*, 165, 61
- Allain, M., Idier, J., & Goussard, Y. 2004, submitted to *IEEE Transactions on Image Processing*
- Bhandari, R. 1978, *Astronomy and Astrophysics*, 70, 331
- Blake, A. & Zisserman, A. 1987, *Visual reconstruction* (Cambridge, MA: The MIT Press)
- Bontekoe, T. R., Koper, E., & Kester, D. J. M. 1994, *Astronomy and Astrophysics*, 284, 1037
- Bouman, C. A. & Sauer, K. D. 1993, *IEEE Transactions on Image Processing*, 2, 296
- Buck, B. & Macaulay, V. A. 1989, in *Maximum Entropy and Bayesian Methods*, ed. J. Skilling (Dordrecht, The Netherlands: Kluwer Academic Publishers)
- Ciuciu, P., Idier, J., & Giovannelli, J.-F. 2001, *IEEE Transactions on Signal Processing*, 49, 2201
- Coulais, A. 1997, PhD thesis, Université de Paris VII
- Daniell, G. J. & Gull, S. F. 1980, *Proceedings of the IEE*, 127E, 170
- Demoment, G. 1989, *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-37, 2024
- Dixon, D. D., Johnson, W. N., Kurfess, J. D., et al. 1996, *Astronomy and Astrophysics Supplement*, 120, 683
- Fomalont, E. B. 1973, *Proc. IEEE*, Special issue on radio and radar astronomy, 61, 1211
- Geman, S. & Geman, D. 1984, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6, 721
- Gull, S. F. 1989, in *Maximum Entropy and Bayesian Methods*, ed. J. Skilling (Dordrecht, The Netherlands: Kluwer Academic Publishers), 53–71
- Gull, S. F. & Daniell, G. J. 1978, *Nature*, 272, 686
- Gull, S. F. & Skilling, J. 1984, *Proceedings of the IEE*, 131-F, 646

- Hogbom. 1974, *Astronomy and Astrophysics Supplement*, 15, 417
- Hunt, B. R. 1977, *IEEE Transactions on Communications*, C-26, 219
- Idier, J., ed. 2001a, *Approche bayésienne pour les problèmes inverses* (Paris: Trait IC2, Srie traitement du signal et de l'image, Hermès)
- Idier, J. 2001b, *IEEE Transactions on Image Processing*, 10, 1001
- Kerdraon, A. & Delouis, J. 1997, in *Coronal Physics from Radio and Space Observations*, 192–+
- Kerdraon, A. & Mercier, C. 1983, *Astronomy and Astrophysics*, 127, 132
- Komesaroff, M., Narayan, R., & Nityananda, R. 1981, *Astronomy and Astrophysics*, 93, 269
- Lannes, Anterrieu, & Marchal. 1997, *Astronomy and Astrophysics Supplement*, 123, 183
- Lantos, P. & Alissandrakis, C. E. 1996, *Solar Physics*, 165, 83
- Le Besnerais, G., Bercher, J.-F., & Demoment, G. 1999, *IEEE Transactions on Information Theory*, 45, 1565
- Macaulay, V. A. & Buck, B. 1985, in *Maximum Entropy and Bayesian Methods*, ed. C. R. Smith & W. T. J. Grandy
- Macaulay, V. A. & Buck, B. 1994, in *Maximum Entropy and Bayesian Methods*, ed. J. Skilling (Dordrecht, The Netherlands: Kluwer Academic Publishers), 59–67
- Magain, P., Courbin, F., & Sohy, S. 1998, *The Astrophysical Journal*, 494, 472
- Mohammad-Djafari, A. & Demoment, G. 1988, *Traitement du Signal*, 5, 235
- Mohammad-Djafari, A., Giovannelli, J.-F., Demoment, G., & Idier, J. 2002, *Int. Journal of Mass Spectrometry*, 215, 175
- Mugnier, L., Fusco, T., & Conan, J.-M. 2004, *Journal of the Optical Society of America*, 21, 1841
- Narayan, R. & Nityananda, R. 1984, in *Indirect Imaging*, ed. J. Roberts, URSI, Australia 1983, 281–290
- Narayan, R. & Nityananda, R. 1986, *Astronomy and Astrophysics*, 24, 127
- Nityananda, R. & Narayan, R. 1982, *Journal of Astrophysics and Astronomy*, 3, 419
- Nocedal, J. & Wright, S. J. 2000, *Numerical Optimization*, Series in Operations Research (New York: Springer Verlag)
- O'Sullivan, J. A. 1995, *IEEE Transactions on Image Processing*, 4, 1258
- Pantin, E. & Starck, J.-L. 1996, *Astronomy and Astrophysics Supplement*, 118, 575
- Phillips, D. L. 1962, *J. Ass. Comput. Mach.*, 9, 84
- Pirzkal, N., Hook, R. N., & Lucy, L. B. 2000, in *ASP Conference Series, Astronomical Data Analysis, Software, and Systems IX*, ed. N. Manset, C. Veillet, & D. Crabtree, Vol. 216, 657–660
- Puetter, R. C. & Yahil, A. 1999, in *ADASS VIII, ASP Conference Series*, Vol. 172, 307–316
- Samson, V., Champagnat, F., & Giovannelli, J.-F. 2003, *Detection of Point Objects with Random Subpixel Location and Unknown Amplitude*, *Applied Optics Special Issue on Image processing for EO sensors*
- Schwartz, U. J. 1978, *Astronomy and Astrophysics*, 65, 345
- Snyder, D. L., Schulz, T. J., & O'Sullivan, J. A. 1992, *IEEE Transactions on Signal Processing*, 40, 1143
- Soussen, C. 2000, Phd thesis, Université de Paris-Sud, Orsay, France
- Starck, J.-L., Bijaoui, A., Lopez, B., & Perrier, C. 1994, *Astronomy and Astrophysics*, 283, 349
- Starck, J.-L., Murtagh, F., Querre, P., & Bonnarel, F. 2001, *Astronomy and Astrophysics*, 368, 730
- Starck, J.-L., Pantin, E., & Murtagh, F. 2002, *Publication of the Astronomical Society of the Pacific*, 114, 1051
- Tarantola, A. 1987, *Inverse problem theory: Methods for data fitting and model parameter estimation* (Amsterdam, The Netherlands: Elsevier Science Publishers)
- Thompson, A. R., Moran, J. M., & Swenson, G. W. J. 2001, *Interferometry and Synthesis in Radio-astronomy* (New York, NY: Wiley (Interscience))
- Tikhonov, A. 1963, *Soviet. Math. Dokl.*, 4, 1624
- Tikhonov, A. & Arsenin, V. 1977, *Solutions of Ill-Posed Problems* (Washington, DC: Winston)
- Twomey, S. 1963, *J. ACM*, 10, 97
- Wakker, B. P. & Schwarz, U. J. 1988, *Astronomy and Astrophysics*, 200, 312
- Weir, N. 1992, in *ASP Conference Series, Astronomical Data Analysis, Software, and Systems I*, ed. D. Worrall, C. Biemesderfer, & J. Barnes, Vol. 25, 186–190